Brit. J. Phil. Sci. **62** (2011), 489–517

A New Argument for Mind–Brain Identity

István Aranyosi

ABSTRACT

In this article, I undertake the tasks: (i) of reconsidering Feigl's notion of a 'nomological dangler' in light of recent discussion about the viability of accommodating phenomenal properties, or qualia, within a physicalist picture of reality; and (ii) of constructing an argument to the effect that nomological danglers, including the way qualia are understood to be related to brain states by contemporary dualists, are extremely unlikely. I offer a probabilistic argument to the effect that merely nomological danglers are extremely unlikely, the only probabilistically coherent candidates being 'anomic danglers' (not even nomically correlated) and 'necessary danglers' (more than merely nomically correlated). After I show, based on similar probabilistic reasoning, that the first disjunct (anomic danglers) is very unlikely, I conclude that the identity thesis is the only remaining candidate for the mental–physical connection. The novelty of the argument is that it brings probabilistic considerations in favor of physicalism, a move that has been neglected in the recent burgeoning literature on the subject.

The mind–brain identity thesis starts its career—setting aside temporally prior and argumentatively and conceptually frugal assertions in that direction by various philosophers and scientists—in the second half of the 1950s, with the work of Ullin Place ([1956]), Herbert Feigl ([1958/1967]), and Jack Smart ([1959]). From today's perspective, Feigl's study, 'The "Mental" and the "Physical" ', is the most remarkable of the three *loci classici* mentioned above in that it is wide ranging, both theoretically and historically, and

doi:10.1093/bjps/axr001

Advance Access published on May 16, 2011

seminal in more than one respect; one can identify in it a large set of topics that are widely discussed today in the philosophy of mind: intentionality, qualia, neural correlates of consciousness, multiple realizability, mental causation, weak and strong reduction, etc.

In this article I undertake the task of reconsidering Feigl's notion of a 'nomological dangler' in light of recent discussion about the viability of accommodating phenomenal properties, or qualia, within a physicalist picture of reality. I will construct an argument to the effect that nomological danglers, including the way qualia are understood to be related to brain states by contemporary dualists, are extremely unlikely. The final step of the argument, the one to the likelihood of the identity thesis, will partially overlap with Feigl's main reason, namely, parsimony, or Ockham's razor. However, unlike Feigl, and those who discussed his idea of nomological danglers, I emphasize that the problems with naturalistic (nomological) dualism are not in the first instance with the 'dangling' bit, but with the 'nomological' bit. I will start with a very brief review of Feigl's above-mentioned notion and the role it plays in his argument for the identity thesis, after which I briefly review the post-Feigl dialectic regarding physicalism and dualism. Then I offer a probabilistic argument to the effect that merely nomological danglers are extremely unlikely, the only probabilistically coherent candidates being 'anomic danglers' (not even nomically correlated) and 'necessary danglers' (more than merely nomically correlated). After I show, based on similar probabilistic reasoning, that the first disjunct (anomic danglers) is very unlikely, I conclude, by the above-mentioned principle of parsimony and two other plausible principles, that the identity thesis is the only extremely likely candidate for the mental–physical connection. The novelty of the argument is that it brings probabilistic considerations in favor of physicalism, a move that has been neglected in the recent mushrooming literature on the subject.

1 The Notion of a Nomological Dangler

Feigl's main objective is to defend the coherence and plausibility of the mind–brain empirical identification thesis, put forward earlier by U.T. Place.[1] He starts by arguing that mental–physical parallelism, i.e. the existence of laws of correlation between mental and physical events or properties, is superior to interactionist dualism. Parallelism is simply the view that there is a law-like connection between the two domains, the mental and the physical, so

[1] As David Armstrong notes (personal communication) 'it was Place who started it all, but unfortunately he published his idea in the wrong place'. The identity thesis then starts its real career with J.J.C. Smart's article, published in 1959, and culminates with Armstrong's ([1968]) and David Lewis's ([1966], [1970], [1972]) functionalist, semantics-based arguments for it.

it is assumed to be a notion neutral between a dualist reading, in which case it is equated with epiphenomenalism, and a physicalist one. Accepting parallelism as plausible is a first step towards the identity thesis. The next step is to observe that these laws, if irreducible, are very different from other lawful generalizations that are present in sciences. Hence Feigl writes:

> These correlation laws are utterly different from any other laws of (physical$_2$) science in that, first, they are nomological 'danglers', i.e., relations which connect intersubjectively confirmable events with events which *ex hypothesi* are in principle not intersubjectively and independently confirmable. Hence, the presence or absence of phenomenal data is not a difference that could conceivably make a difference in the confirmatory physical$_1$-observational evidence, i.e., in the publicly observable behavior, or for that matter in the neural processes observed or inferred by the neurophysiologists. And second, these correlation laws would, unlike other correlation laws in the natural sciences, be (again ex hypothesi) absolutely underivable from the premises of even the most inclusive and enriched set of postulates of any future theoretical physics or biology.' ([1958/1967], Chapter 5, Section B, p. 61)

Feigl defines two notions of the physical in Chapter 5, Section A. 'Physical$_2$' is defined as the kind of theoretical concepts and statements which are sufficient for the explanation of the observation statements regarding the inorganic (lifeless) domain of nature, whereas 'physical$_1$' refers to the concepts and statements used by all sciences, which involves logical or probabilistic connections to intersubjective observation language. As a matter of fact, according to Feigl, the two domains of the physical actually coincide, which means that there are no genuinely emergent properties.

Turning now to the notion of nomological danglers, we observe that for Feigl it applies to certain nomic relations, and not to the relata of these relations.[2] So the 'nomological' part of the notion refers to the fact that psycho-physical correlations represent, indeed, some form of nomic connection. However, the 'dangling' part refers to the fact that, unlike other scientific laws, psycho-physical ones are odd in that they relate the standard, publicly observable, and intersubjectively confirmable phenomena (the brain states) with phenomena that are *ex hypothesi* exclusively subjective (the raw feels, or qualia), and hence they don't (and can't) make any explanatory difference when it comes to confirmation of a hypothesis about a potential nomic connection.[3]

[2] Unlike for Smart ([1959], p. 142), who adopts the phrase from Feigl, but changes its meaning so as to apply to sensations, raw feels, or qualia, which are supposed to dangle, i.e. to be ontologically distinct from but lawfully connected to what the complete scientific picture of the world would encompass.

[3] Cf. David Chalmers' ([1996]), where he uses, throughout his book, the formula 'the explanatory irrelevance of conscious experience' to express the same thought.

The notion of a nomological dangler is not logically incoherent. The only problem with it, according to both Feigl and Smart, is the 'dangling' bit, but that is not a problem of logical coherence; it is rather a problem of suitability to the naturalistic and reductionist *Zeitgeist*. That being said, Feigl completes his argument for the identity thesis by arguing that once the parallelist or correlationist thesis is accepted as plausible, considerations of ontological parsimony and the methodological constraint of avoiding nomological danglers as unnecessary make the empirical (i.e. non-analytic) identification of mental and neurophysiological types of states the only plausible candidate for the mind–body relation. Once the identification has been made, the way the mental is related to the physical is no different from the way particular aspects of the physical are related to the physical, and these nomic connections are between relata that are intersubjectively available and confirmable.

Following the great opening due to the three above-mentioned materialist thinkers, the empirical identity thesis was turned, in the second half of the 1960s, in the works of David Lewis ([1966]) and David Armstrong ([1968]), into a thesis with an essential analytic component. The reason Feigl had been afraid of analyticity in connection with the mind–body problem was that he couldn't see, and rightly so, any prospect for synonymy between the neurophysiological and the phenomenal vocabulary. However, it was actually Smart who first realized[4] that even if there is no direct synonymy between these vocabularies, one could formulate an analysis, or at least a gloss on phenomenal concepts by using a so-called topic-neutral vocabulary:[5]

> My suggestion is as follows. When a person says, 'I see a yellowish-orange after-image,' he is saying something like this: '*There is something going on which is like what is going on when* I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange.'[. . .]. Notice that the italicized words, namely 'there is something going on which is like what is going on when,' are all quasi-logical or topic-neutral words. This explains why the ancient Greek peasant's reports about his sensations can be neutral between dualistic metaphysics or my materialistic metaphysics. (Smart [1959], pp. 149–50)

[4] In fairness to Feigl, we should mention that he comes very close to the identity theory as mediated by topic-neutral analysis, as he actually uses a two-step identification process, one between raw feels and referents of concepts that are inferentially related to logically behavioral concepts, and one from the latter to referents of neurophysiological concepts. At the very beginning of Section E of Chapter 5 (p. 78) he writes: 'Taking into consideration everything we have said so far about the scientific and the philosophical aspects of the mind-body problem, the following view suggests itself: The raw feels of direct experience as we "have" them, are empirically identifiable with the referents of certain specifiable concepts of molar behavior theory, and these in turn [. . .] are empirically identifiable with the referents of some neurophysiological concepts.'

[5] Indeed, as Colin McGinn ([2001], p. 286) points out, one finds two essentially different theories in one and the same article by Smart under the name 'identity theory': what has nowadays been called 'a posteriori' versus 'a priori physicalism' (cf. Stoljar [2000], [2001]).

The topic-neutral analysis would later become the standard step towards the identification of mental and neurophysiological properties in the doctrine of analytic functionalism. First, mental terms are given an analysis in causal-role functional terms, then the realizer of the causal role in actuality is identified as a certain neurophysiological property. Thus, the identity thesis is argued for in a way that does not appeal to Ockham's razor (cf. Lewis [1966]).

The 1970s and 1980s have witnessed the emergence of a series of important anti-physicalist arguments.[6] I am not going to expound these arguments, but rather rely on Chalmers' approach to them and synthesize them under the general heading of epistemic arguments. According to Chalmers ([2003], p. 108):

> The general form of an epistemic argument against materialism is as follows:
>
> (1) There is an epistemic gap between physical and phenomenal truths.
>
> (2) If there is an epistemic gap between physical and phenomenal truths, then there is an ontological gap, and materialism is false.
>
> ———————————
>
> (3) Materialism is false.
>
> Of course this way of looking at things oversimplifies matters, and abstracts away from the differences between the arguments. [. . .] Nevertheless, this analysis provides a useful lens through which to see what the arguments have in common, and through which to analyse various responses to the arguments.

The epistemic arguments apply, of course, not only to physical properties as such, but the topic-neutral ones that are used by analytic versions of the identity thesis. It is argued in the first premise that the instantiation of both physical and functional properties is epistemically compatible with the lack of instantiation of phenomenal properties, or raw feels. Then it is inferred that since there is no reason to doubt in this particular case that the epistemic compatibility is a good guide to metaphysical compatibility, the corresponding metaphysical compatibility claim is justified. Hence, physicalism—the view that the totality of actually instantiated physical and functional properties metaphysically entails the instantiation of raw feels—must be false.

If these epistemic arguments are accepted as sound, then if we combine them with Feigl's idea that nomological danglers, though methodologically weird, are logically coherent, we get the doctrine of naturalistic dualism as a

[6] Saul Kripke's argument against the early, empirical, and contingent identity thesis ([1972]), the argument from the conceivability of zombies (Kirk [1974a], [1974b], and later revived and developed by Chalmers [1996]), the argument from subjectivity (Nagel [1974]—though we should add that Nagel himself did not take his argument to actually prove the falsity of physicalism), the argument from the explanatory gap (Levine [1983]—as in Nagel's case, we should note that Levine did not think his argument was incompatible with the truth of physicalism), the knowledge argument (Jackson [1982]).

perfectly coherent and plausible view of the mind–body relation. According to naturalistic dualism, the mental and the physical properties are both fundamental to the actual world in the sense that neither of them metaphysically supervenes on the other. They are ontologically distinct kinds of properties. Nevertheless the 'naturalistic' bit of naturalistic dualism asserts that the two kinds of properties are nomically connected, namely, by laws of nature whose form we can infer by extending our own case of the link between our raw feels and functional-cum-neurophysiological properties. Finally, these nomic connections are contingent, just like other laws of nature, according to this doctrine. The view[7] is based on the apparently unproblematic observation that there is nothing incoherent in the idea that phenomenal properties figure in special, irreducible, and fundamental psycho-physical laws. As we have seen, Feigl himself, although uneasy about the 'aesthetics' and skeptical about the *indispensability* of such laws, did not think they are incoherent as such.

This brings me to the argument I would like to put forward, to the conclusion that the doctrine of naturalistic dualism is probabilistically incoherent, and that physicalism, in the form of the identity thesis, is the likeliest candidate for the mental–physical relation.

2 The Probabilistic Incoherence of Naturalistic Dualism

The idea of nomological danglers, as understood by Feigl, involves the idea of asserting the existence of a nomic connection between the intersubjectively confirmable phenomena and some phenomena that are in principle unavailable intersubjectively. How do we know that a certain nomic pattern holds in the actual world, if this is so? One thing that both Feigl and recent naturalistic dualists accept is that we can't be sure of it, but they also agree that we can probabilistically infer, from behavior and other intersubjectively available evidential bases, the existence of the relata of this connection, and therefore we can infer the existence of the supposed nomic connection. But this is too swift. Even if we can probabilistically infer the existence of subjective raw feels from intersubjective data and own case subjective data, that is still not enough to infer that the correlations that hold in my own case—of the form 'whenever I am in intense pain, I am screaming', etc.—hold in exactly the same way for others. Nevertheless, suppose we do have evidence for probabilistically inferring the existence of the nomic connections the way we think they are.

Let us call the psycho-physical nomic profile of the actual world that universally generalizes own case phenomenal–neurophysiological correlations in

[7] Among its supporters we find Chalmers ([1995], [1996]), Tim Crane ([2001]), Galen Strawson ([1997]), Leopold Stubenberg ([1998]), and myself ([2008]).

the relevant subjects 'the normal nomic profile' (NNP). If the actual world is in NNP, then it is always the case that whenever a subject sees red, she is disposed to assert 'I see red', whenever a subject feels pain, she is disposed to manifest it behaviorally by exclaiming 'Ouch!', and so on and so forth. If these are the correlation laws, then NNP is how the naturalistic dualist conceives of the actual world. There are no spectrum-inverted pairs of subjects, there are no zombies, etc. As I said, *ex hypothesi*, we can never be sure whether others are undergoing the very same phenomenal states as we do when the same behavioral and neurophysiological properties are instantiated in them. We supposed that, nevertheless, we have probabilistic justification to think that the world is in NNP. The problem with naturalistic dualism is that the very same premise that ensures the logical coherence of dualism, namely, the first premise of the canonical epistemic argument, also weakens to almost null degree the probabilistic justification for the proposition that NNP is actually the case. So if that premise is plausible, then our probabilistic inference to NNP in the actual world is extremely implausible or flawed.

Let me explain. The first premise of the epistemic argument asserts that there is an epistemic gap between physical and phenomenal facts. Translated in the form of a conceivability claim the premise would assert that it is conceivable that there be a world that is physically exactly like the actual world, but phenomenally different. The formula 'phenomenally different' is very general, indeed, so the number of conceivable scenarios of combinations of phenomenal properties is, on the assumption of discreteness of these properties and the conceivability of alien properties,[8] countably infinite. Furthermore, there is a sense in which (see Section 4 for an account of this claim) the number of anomic scenarios will be by far greater than the number of conceivable nomic profiles, as it will be a function of combining instantiations of phenomenal properties even at the level of one particular subject, or one particular time. Of course, the extant literature on conceivability arguments got us conditioned on a couple of rhetorically salient cases, true Schelling points of the logical space, namely the zombie scenario (when everything is physically as it actually is, but there is no instantiation of any phenomenal property whatsoever), and the inverted qualia scenario (when everything is physically as it actually is, but color qualia instantiations are spectrum inverted with respect to the actual world). When discussing conceivability arguments, we typically focus on these scenarios. But we shouldn't. They are purely rhetorical devices to make the anti-physicalist argument

[8] Alien properties are defined as those that are instantiated in some possible worlds, but not in the actual world.

appear intuitive. If zombies and qualia inversion are conceivable, then there is no principled reason for all the other, infinitely many scenarios not to be conceivable. This brings us to formulating what I would call 'the principle of explosion'.[9]

> (EXPLOSION) If a scenario *S* is conceivable, then all relevantly similar scenarios are conceivable.

Of course, it is not always the case that we can straightforwardly find out whether a pair of scenarios are relevantly similar, but in our case—the psycho-physical case—the similarity is underwritten by (i) physical duplication and (ii) phenomenal difference. It is because phenomenal difference involves any conceivable combination of phenomenal properties that we are able to apply the principle to the psycho-physical case:

> (PSYCHO-EXPLOSION) If a physical duplicate of actuality that is phenomenally different in respect *R* is conceivable, then all physical duplicates of actuality that are phenomenally different in any respect *R** are conceivable.

But if there are infinitely many conceivable non-NNP scenarios, most of which are anomic, i.e. random distributions of phenomenal properties in physical duplicates of our world, how are we to know, probabilistically, that our world has NNP? We know from the discussion on Feigl that phenomenal properties are not intersubjectively accessible. Similarly, Chalmers argues at length for the explanatory irrelevance of phenomenal properties, which means that they do not make a difference to third-person observation, and, hence, to evidential bases for confirmation. But this means that for all we know intersubjectively, the actual world could turn out to be in any conceivable nomic or anomic profile. The principle of indifference, which is a rule for assigning probabilities under ignorance, would tell us that the actual world being in NNP is equiprobable with any of the infinite number of other conceivable scenarios. That means that the probability of the actual world being in NNP is $1/\infty$, i.e. zero. However, by contraposition, since the naturalistic dualist believes that the actual world is surely in NNP, it means that the non-NNP scenarios (zombies, inversions, etc.) must be inconceivable—they are not present in logical space.

[9] The principle is inspired by work on impossible worlds. An instance of explosion is the *ex falsum sequitur quodlibet* in standard logic, according to which one can derive any proposition from any contradiction. It implies that, in a world where one contradiciton is true, everything is true. I borrow the term 'explosion' from Daniel Nolan ([1997]) who calls a world where every proposition is true an 'explosion world'. Nolan argues against the principle of explosion, Lewis ([1988]) offers an argument for it.

3 The Inference to Mind–Brain Identity

I would like to go further and argue for a way to reach, at this point, the step of identifying the mental and the physical, just like the identity theory would prescribe. There are two ways to go, both of which seem to me quite plausible.

One way is to assume naturalism in the form of accepting that the actual world has NNP, and, consequently, to deny the disjunct stating that the actual world is very likely to be anomic with respect to psycho-physical connections. This would entail the inconceivability of zombies, qualia inversion, and other abnormal scenarios, and consequently the epistemic necessity of the actual psycho-physical nomic profile. The options we are then left with are: (i) distinctness but necessary connection between phenomenal and physical properties; and (ii) identity.

There are three arguments against the first option. One argument is via Hume's dictum that there are no necessary connections between distinct existences. Contraposing, since we have reason to think that the necessary connection must hold, we have reasons to identify the phenomenal with the physical. Another argument is via Ockham's razor, which states that one shouldn't multiply entities beyond necessity. In other words, if one can explain some phenomenon in various ways, one should opt for the ontologically most parsimonious such way. Applied to our case the argument would make plausible the identity thesis in comparison to the less parsimonious alternative based on necessary connections between distinct kinds of properties, given that both options have exactly the same explanatory status and make the same predictions. Finally, the third argument is from the principle of no brute, unexplained necessity, which states that postulated necessities should always be grounded in logical necessity. The postulated necessary nomic profile is definitely a brute necessity, but if the identity thesis is adopted, then the relevant necessity follows from the logical property of the necessity of identity.[10] Chalmers, for instance, reproaches certain versions of physicalism for not explaining their necessity claim. If my argument is right, then the physicalist necessity is explicable via a priori probabilistic reasoning, while the dualist response in terms of necessary pyscho-physical laws comes out as involving brute modal facts.

[10] One might ask at this point: what explains the identity itself? David Papineau ([2002], p. 114) argues that identity, in general, is in no need of explanation; it does not make good sense to ask, once we know an identity to hold, why that identity holds. In our context, however, we need not rely on such a principle. What explains our mind–brain identity is precisely the fact that that identity itself explains our thesis of necessary correlation between mental and physical properties. It is not infrequent in science that our commitment to the existence of some x is explained by the fact that x explains, in the best available way, some y. For example, the commitment to the existence of the gene is explained by the fact that the gene explains, in the best available way, our observations about heritability of traits.

Another way to arrive at the step of identification is not by *assuming* naturalism, i.e. the conformity of the actual world with NNP, but by weighing the two conflicting reasons, the one for naturalism (and against the conceivability intuition) and the one for the conceivability of an infinity of physical duplicates with various combinations of phenomenal property distribution. If the reasons for naturalism are stronger, then we can eliminate the anomic scenario and go through the three above-mentioned arguments to the conclusion that the identification of the mental with the physical is to be preferred. The question is then which of the reasons is stronger. My argument for the necessity of the actual NNP and against the conceivability of zombies, qualia inversion, and other combinations of phenomenal property distributions relies on the fact that the reasoning based on the indifference principle can safely be taken as simply an extension of the a priori reasoning involved in thoroughly considering the conceivability of the zombie and other non-NNP scenarios; hence, since the end result of this a priori reasoning is something that is incompatible with the initial conceivability intuition, we should discard that intuition. It is only *prima facie* conceivable that there be zombies and other non-NNP scenarios. We may start by applying the indifference principle to our own phenomenal states (which is directly evidential) with the reference class of actual observers, and thereby establish that the actual world has to conform, almost surely, to NNP (see more about this in Section 4). Then we apply the indifference principle once again, with the actual world taken as a random sample of the set of all initially conceivable worlds that are physical duplicates of the actual world but phenomenally different. Then we conclude that since otherwise our world conforming to NNP would be a huge coincidence, it must be the case that it is sure that all physical duplicates of our world are to be assumed as being mental duplicates as well, which is equivalent to concluding that our initial conceivability intuition was wrong and has to be revised in light of the subsequent a priori probabilistic reasoning.

4 The Technical Formulation of the Argument

Mathematically, given that the conditions for the applicability of the indifference principle are satisfied, the probability of the actual world, @, being in NNP—$p(@_{nnp})$—is given by the fraction (f_{nnp}) of NNP-worlds (N_{nnp}) within the total number of physical duplicates of the actual world ($N_{nnp} + N_{non\text{-}nnp}$):

$$p(@_{nnp}) = f_{nnp} = N_{nnp}/(N_{nnp}+N_{non-nnp})$$

We observe that $p(@_{nnp})$ approaches unity (i.e. our belief that the actual world is in NNP is almost certain) when, and only when the difference between

($N_{nnp} + N_{non\text{-}nnp}$) and N_{nnp} approaches zero, that is, when $N_{non\text{-}nnp}$ is approximately zero. Accepting that $N_{non\text{-}nnp}$ is approximately zero is tantamount to denying either the conceivability premise, or PSYCHO-EXPLOSION. Since the latter is undeniable by the naturalistic dualist, she is then forced to choose between $p(@_{nnp})$ being zero and the denial of the conceivability premise.[11]

Now, in the formula above I assumed $N_{nnp} + N_{non\text{-}nnp}$ merely to approach infinity. However, as I pointed out before, we can take, under some assumptions, the number of possible phenomenal configurations consistent with physical duplicates of @ to be countably infinite. In that case, if both PSYCHO EXPLOSION and the thesis that psycho-physical correlations are nomological danglers, in Feigl's sense, are true, then the conceivability of some non-NNP scenario entails that the actual world being in NNP has probability zero (assuming a standard infinite probability space), or infinitesimal (assuming a nonstandard, i.e. hyperreal-valued, probability space). One intuitive difference between the standard setting and the nonstandard one is that the former will allow for possibilities with probability zero, whereas the latter will reserve probability zero for logical impossibilities only. Hence, nonstandard spaces, based on the hyperreal line, are *prima facie* more intuitive from

[11] We get similar results in a Bayesian framework. Let's denote by H the hypothesis that there is a huge number of physical duplicates of @ that differ in phenomenal property distributions, with E the proposition that @ is in NNP, and with ¬H the proposition that almost all physical duplicates of @ are in NNP.
Bayes Rule says: $P(H|E) = p(H)p(E|H)/ [p(H)p(E|H) + p(\neg H)p(E|\neg H)]$. Let us assign some very high probability to H, say, .99999. By the indifference principle we get a very low probability for p(E|H), and a very high one, i.e. 1 – p(E|H), for p(E|¬H). The numerical values for our parameters are as follows:

H: there are infinitely many conceivable phenomenal distributions over physical duplicates of @
p(H) = .99999

¬H: almost all physical duplicates of @ are in NNP
p(¬H) = .00001

E: @ is in NNP

p(E|H) = .0000001

p(E|¬H) = .9999999

Effecting the replacement in Bayes Rule, we get:

$P(H|E) = .99999*.0000001/[.99999*.0000001 + .00001*.9999999]$

$=9.9999*10^{-8}/(9.9999*10^{-8} + 9.999999*10^{-6}) = .0099$

This means that one should update p(H) from .99999 to .0099. If the number of non-NNP physical duplicates of @ approaches infinity, then the posterior probability of H is approximately zero.

this and other points of view[12] (Skyrms [1995]; Vallentyne [2000]; Elga [2004]; Herzberg [2007]; Williamson [2007]).[13]

However, in both standard and nonstandard settings it will be true that if the probability of @ being in NNP is strictly positive noninfinitesimal (i.e. nonzero in standard setting and noninfinitesimal in nonstandard setting), then the event '@ is in NNP' has to be sure (i.e. '@ is in NNP' is an atom[14] of probability 1 in the sample space) rather than almost sure (i.e. the sample space has either a union of probability 0 of non-NNP states, or a union of infinitesimal probability of non-NNP states). This is so because the thesis of nomological dangling ensures that all the events of the sample space have to be equiprobable; both the hypothesis of a sample space with a union of probability 0 of non-NNP states, and one with a union of infinitesimal probability of non-NNP states would contradict the requirement of equiprobability of all events. To say that @ is in NNP surely is, therefore, to say that it is the only possibility. Hence, whenever one assigns a higher than infinitesimal probability to NNP being actual, one has to admit that all non-NNP scenarios are epistemically impossible; hence, the actual world is in NNP *surely*. Contraposing, whenever one assumes that any non-NNP scenario is epistemically possible, one has to admit that the actual world being in NNP has probability zero or infinitesimal, so the actual world is not in NNP *almost surely*.[15]

[12] For instance, in standard probability theory, because any atom of a countable infinite state space has measure zero, any subset of the union of such atoms has the same probability as the union (since they have the same cardinality), assuming uniform probability distribution over the infinite number of states. Also, any finite union of atoms has the same probability as an infinite subset of the state space. To take an example for each, if we are to choose a number randomly from the set of natural numbers, the event 'the number is a multiple of 2' and 'the number is a multiple of 100' have the same probability, and the same is true of events 'the number is between 1 and 1 million' and 'the number is a multiple of 3'. Some philosophers find these facts counter-intuitive (e.g. McCall and Armstrong [1989] and Vallentyne [2000]). Things are different with nonstandard probability theory, as I exemplify later, in footnote 16.

[13] We should note, though, that all these authors except Vallentyne and Herzberg have arguments that discourage a too optimistic attitude towards nonstandard analysis as a more intuitive basis for probability theory.

[14] An atom of a probability space is a set of strictly positive measure such that any measurable subset of it has either that measure or measure zero. A probability space with an atom of probability 1 is called 'trivial'. A probability space $(\Omega, \mathbf{A}, P)$, where Ω is the sample space, $\mathbf{A}$ is an algebra on Ω, and P is the probability function, is trivial iff we only have $\varnothing$ and Ω as events in it, that is, $\mathbf{A} = \{\varnothing, \Omega\}$, it consists of exactly two sets—the sample space (everything) and the empty set (nothing).

[15] The argument works for a continuous probability space just as well. Assume that the sample space is the real unit interval and our variable, the phenomenal space, is a continuous random variable. This means that instead of probabilities as such we will have a probability density function, whose integral over an interval of possible values will assign a (a nonzero) probability for the actual world being within that interval. If any non-NNP world is conceivable, then by PSYCHO-EXPLOSION, continuously many of them are conceivable. Given the thesis of nomological dangling applied now to phenomenal continua, the indifference principle applies across all values of the phenomenal variable across physical duplicates of @. Hence, the probability of @ being within some interval $(a, b]$, such that $0 \leq a < b \leq 1$, corresponding to NNP, is given by the continuous uniform distribution, and will have to be the same as that of @ being in

Furthermore, if we adopt nonstandard analysis, there is a sense in which the number of anomic distributions of phenomenal properties across physical duplicates of the actual world is by far larger than nomic distributions, and we also get the disturbing result that it is almost certain that we live in a world with a random distribution of phenomenal properties. Here, I use 'nomic' as a qualifier of a distribution of phenomenal properties over a physical duplicate W of @ to denote a (contingent) supervenience preserving pattern of physical–phenomenal coinstantiations (so for each world W, it is a surjection from the domain of actually instantiated properties to a codomain consisting of phenomenal property instantiations in W). An 'anomic' distribution will be one that does not have this structure. For instance, a world in which each type of brain state B is always associated with some particular type of phenomenal state A is one in which A has a nomic distribution in the above sense, whereas a world in which B is associated with different types of phenomenal states at different times or places is one in which the phenomenal distribution is considered anomic. The claim is that such anomic distributions are by far more probable than nomic ones, in nonstandard settings.

We don't know the exact pattern of physical property instantiations of the actual world, and even if we did it would be too complicated to be used to exemplify this point. So let us use a toy model of the physical aspect of the actual world. Let us assume that the actual world exists for two moments of time, and a brain state B is instantiated both times. Supervenience preserving combinations of phenomenal property instantiations (i.e. what I have called 'nomic distributions') will be sequences of two terms, a_1 and a_2, such that $a_1 = a_2$. For instance, if we represent the possible phenomenal states as the set of natural numbers, the nomic distributions in our toy model will be {1,1}, {2,2}, . . . We are interested in the proportion of such distributions within the set of all possible 2-permutations of the set of phenomenal properties. In standard probability theory this number would be the same as the total number of possible 2-permutations, as the two sets have the same cardinality (that of $\mathbb{N}$). But in nonstandard analysis the ordinary algebraic operations can be applied to nonstandard infinities (also called 'unlimited hyperreals') on the model of the finite case[16]. We denote the nonstandard infinite number of

any interval of equal length (measure), i.e. length $a - b$. If that is the case, then as $a - b$ approaches zero the probability of @ being in NNP either approaches a value that is statistically equivalent to zero (in measure-theoretic terms: the property of being non-NNP holds *almost everywhere*); or otherwise it is strictly greater than zero, in which case $(a, b] = \Omega$ almost everywhere, which means that the only events in the sample space are NNP and the null set (so the non-NNP scenarios do not exist as epistemic possibilities at all).

[16] The reader might want to consult Jerome Keisler's ([2000]) introduction to nonstandard analysis for the properties of algebraic operations having hyperreal numbers as their terms. I note here only that an infinitesimal (or infinitely small) number is a ε, such that $-a < \varepsilon < a$, for all positive real numbers a. The only real number that is infinitesimal is zero. The line of hyperreal numbers

sequences of the form $\{a_1, a_2\}$, such that $a_1 = a_2$, with H. The total number of sequences, i.e. distributions of phenomenal properties, in our toy model, obtained by applying the formula for n-permutations of a set of x elements with repetitions,[17] is then $H + H!/(H-2)!$. The proportion of H in this set is (after applying the rules of division for factorials) H/H^2, which is infinitesimal (Keisler [2000], p. 32). This means that it is infinitely more probable that the actual world is anomic in the sense of phenomenal property instantiations not conforming to the requirement of intra-world supervenience.

However, the indifference principle is of help here once more. The principle, in its most general form, asserts that when there is no a priori reason to assign more probability to an outcome than to any other outcome, one should assign equal probabilities (or degrees of belief) to all possible outcomes. A version of this principle is formulated by appeal to observers and a reference class. One such version, due to Nick Bostrom ([2002a]) is the self-sampling assumption:

> (SSA) One should reason as if one were a random sample from the set of all observers in one's reference class.

As we noted before, Feigl argued that we can infer, probabilistically, from our own case that our world conforms to NNP. Naturalistic dualists agree, as they think the alternative would be blanket skepticism about other minds, and the unacceptable conclusion that even if there are other minds these are randomly related to physical states. We can argue for such an inference from own case to NNP being actual by appeal to SSA. What I observe is that I myself do conform to what NNP would predict an observer should observe. I observe that my behavior and my brain states always match my phenomenal states.[18]

is then constructed by positing infinitesimals that are not zero and adding them to the real line. A positive nonstandard infinite number, H, is then $1/\varepsilon$. Standard algebraic operations and relations can be applied in this setting so that we get: negatives, reciprocals, sums, products, quotients, and roots. To give a few examples, $1/H$, ε/H, and $\varepsilon/1$ are infinitesimals; $H/1$, H/ε, and $1/\varepsilon$ are infinite (provided that $\varepsilon \neq 0$); H/K, ε/δ, $H\varepsilon$, and $H + K$ are indeterminate forms, their value depending on what H, K, ε, and δ are. For instance, if ε is $1/H$, then $H\varepsilon = H/H = 1$; if ε is $1/H^2$, then $H\varepsilon = H/H^2 = \delta$ (an infinitesimal). See (Keisler [2000], Chapter 1).

[17] The number of n-permutations of a set of x elements *without* repetitions is $x!/(x-n)!$, which is equivalent to the number of injective functions from x to n. Adding x to this number we get the number of n-permutations of a set with x elements with repetitions.

[18] An anonymous referee objects that in order to assert that I observe that my own phenomenal states match my physical states as prescribed by NNP presupposes perceptual realism, which in turn depends upon acceptance that the world exemplifies NNP. Also, he/she objects that 'I have zero evidence about my brain states and this is true of almost everyone (I know a few people who have been in MRI experiments etc. but they are very few and far between)'. Regarding the second objection, what is important is not whether I really have such evidence, but that, in principle, I can have such evidence. For instance, I can take images via computer tomography of my brain while having a certain kind of experience, and establish correlations. Regarding the first objection, to insist on having to solve some metaphysical problems in the philosophy of perception in order to be justified in asserting that some phenomenal–neural or phenomenal–behavioral correlations hold would, in general, and from the point of view of the empirical

Given this, what is the probability that all other actual observers undergo the same pattern of psycho-physical correlations? By SSA, I should take myself as a random sample from the set of actual observers. Given what I observe in my own case, the set of actual observers has to contain a much larger number of NNP-conform observers than ones with random phenomenal distributions. That means that it should be almost certain that the actual world has NNP.

One might ask at this point: What if we considered the set of all conceivable observers in all physical duplicates of the actual world? Would we have obtained the same result as previously when considering worlds, namely, the result that it is almost certain that the actual world does *not* have NNP?

First, let us note that the answer is no: had we considered all conceivable observers, the fact that my subjective case does conform to NNP would have been a huge coincidence, unless most conceivable observers conformed to NNP. In order for my experiences to be as they actually are, their probability has to be very high (given that I am one case out of an infinite number of observers), which means, in light of the indifference principle, that almost all conceivable observers have to have experiences conforming to NNP.

Second, the reason for choosing as the reference class the class of actual observers is straightforward: we were interested in whether other people *in the actual world* undergo the same experiences correlated with the same brain states as I do. The reference class had already been selected via the fact that in my own case I do have direct evidence about the correlations. When, on the other hand, we inquired about whether the actual world as such conforms to NNP, we proceeded from ignorance about what probability to assign to each conceivable world; therefore, it was whole worlds that were rightly considered as members of the reference class.

Where does all this leave us? We apparently have conflicting results. By PSYCHO-EXPLOSION, and the first premise of the conceivability argument, we have reason to believe in an infinity of non-NNP worlds, and hence

method, be to set the bar too high, and even to change the subject in a sense. It would mean to set the bar too high because if it were right, then asserting any law, including purely physical ones, would be problematic just because it hasn't been settled whether perception involves the world or some intermediary 'veil' between us and the world, or just sense data. It would mean to change the subject because one can take any view about the metaphysics of perception and accommodate assertions about nomic connections by employing the terminology of that view. For instance, one can be a phenomenalist (the view that reality is constituted by sense data) and formulate laws in terms of phenomenal states that present themselves as physical (as being about physical, chemical, behavioral, etc., facts) and phenomenal states that present themselves as phenomenal (experiences, qualia). One can be a so-called representative realist (the view that experience is constituted by sense data, which in turn represent—by correlation or isomoprhism—a physical reality that lies beyond the 'veil of perception') and formulate laws in terms of correlations between sets or structures of phenomenal property instantiations and the physical structures represented by sense data with a physical content. In this article I have used a realist terminology as that seems the simplest and most intuitive, but nothing hinges on this choice as far as the main argument is concerned.

to believe to a very high degree that our world is anomic in terms of psycho-physical connections. On the other hand, by SSN as applied to our own experience, which conforms to NNP, we should believe to a very high degree that our world does conform to NNP. But if it is almost sure that the actual world has NNP, then, by the first application of the indifference principle, it must be sure that it has NNP, which means that physical duplicates of the actual world with different patterns of phenomenal property distribution are not possibilities at all, they are inconceivable.

Even this much, however, is already damaging for naturalistic dualism. On the assumption that PSYCHO-EXPLOSION is indeed very plausible, we are left with a disjunctive conclusion which excludes, probabilistically, the view that there are psycho-physical laws that are merely nomological danglers:

> (C) Either it is almost sure that our world is psycho-physically anomic, or physical duplicates of the actual world with different patterns of phenomenal property distribution are inconceivable.

To say that these non-NNP scenarios are inconceivable is tantamount to saying that NNP is epistemically necessary. So the two options we are left with are: anomic danglers and logically necessary danglers, but not merely nomological danglers, contrary to the naturalistic dualist doctrine.

Before concluding I would like to consider a few objections. Some of them are related to the core argument, some to technicalities connected to probability theory.

5 Objections Related to the Core Argument

> ***Objection 1:*** *The argument aims at establishing ontological conclusions supposed to be derived from epistemic premises, but one could be skeptical about whether such inferences, in general, are acceptable.*[19]

In response I would like to point out that we have two groups of arguments: ones leading to the partial conclusion expressed by disjunction (C) and then to the assertion of the second disjunct of (C), i.e. that NNP is necessary, and ones for the plausibility of identity as the relation between mental and physical properties. Within the first group, the argument for (C) is purely epistemic with (C) itself being an epistemic claim, or a claim about what it is rational to believe. Then, the second disjunct of (C), which states that non-NNP scenarios are inconceivable, entails the ontological conclusion that these scenarios are impossible, hence another ontological conclusion that physicalism is true. Virtually everyone in the debate agrees that the move from inconceivability to impossibility is not problematic; the disagreement is about whether

[19] An objection put forward by two anonymous referees.

conceivability entails possibility. There is no such claim as the latter in my argument.

As regards the second group of arguments, the ones for the plausibility of identity, these are, again, based on standardly accepted principles: Hume's dictum, the thesis of no brute necessity, and Ockham's razor.

Therefore, there does not seem to emerge any especially worrying issue related to the epistemic-ontic inference.

> ***Objection 2:*** *The cross-world application of SSA is dubious, as its type of conclusion, namely, the necessary truth of the actual nomic profile, would overgeneralize to cases in which, intuitively, actual facts (laws, constants, magnitudes, probabilistic correlations) are to be taken as contingent.*[20]

Take, for instance, the fine-structure constant, which characterizes one of the four fundamental forces, namely, the electromagnetic force, and is responsible, among other things, for our observing a stable chemical structure of the world. This stability is explained, therefore, by a correlation between atomic structure and chemical properties. But consider all possible worlds that are all-law duplicates of the actual world save for the fine structure constant value. A random observer across all these worlds will not observe the correlation between atomic and chemical properties. Do we infer that the fine structure constant necessarily takes on the same value in all the as-defined worlds? Not at all. The fact that the fine structure constant varies across possible worlds does not show that it probably varies within the actual world.

There are two independent answers to this problem. The first one stresses the difference between (i) a case like the one above, involving an established set of laws based on actual observation and (ii) the case of mental–physical correlations in the context of Feigl's point about nomological danglers, or Chalmers's equivalent thesis of the explanatory irrelevance of experience. The difference that is worth stressing here consists of the fact that the issue of whether there is a nomic relation between the relevant terms in case (i) is settled by observation. Similarly, any actual physical law is established, or at least confirmed, by observation. The picture is different with case (ii). Here, because the thesis of the explanatory irrelevance of phenomenal experience is assumed (which states, basically, that since experiences are not intersubjectively available, whether experience occurs in a subject, other than the first-person, does not make a difference when it comes to explaining or predicting the occurrence of intersubjectively available events, e.g. brain states and behavior), the issue is precisely whether it is legitimate to posit a nomic relation on the model of the one identifiable in one's own case (i.e. at the first-person level). The issue cannot be settled, in other words, by observation,

[20] I am grateful to an anonymous referee who has made all the objections that will occur under this heading.

because there is no observation having as content the phenomenal states of others.

In case (i) establishing the nomic connection is independent of considerations about what is conceivable or not—it is simply a matter of what is observed via standard empirical methods. Because of this independence, it does not follow from the fact that there is such an observed nomic connection that it is necessary, i.e. that it is the same in all possible/conceivable worlds. However, in case (ii) establishing the nomic connection is not independent of what is conceivable, because we can't make the relevant third-person observations. Since all random observers across all possible worlds that are physically like ours are in the very same situation as we are with respect to phenomenal experiences from a third-person point of view, the argument goes through, namely, it establishes as a condition on warrantedly asserting that NNP is actually exemplified that this nomic profile is exemplified in all possible worlds. Things would have been different if we had a so-called 'consciousness meter'.[21] Then we would have had both empirical evidence of the actuality of NNP and a priori evidence of its contingency.

The second answer involves the idea that under certain assumptions of current, observationally well-grounded theories in cosmology we do in fact have reason to posit variation within the actual world as a function of cross-world variations regarding various magnitudes and constants. Both multiverse theories (according to which there is an indefinite number of parallel and causally disconnected universes, ours being one of them), and the Big Bang theory combined with the hypothesis that our universe is flat, i.e. it has a Euclidean topology, have the consequence that all possible/conceivable observations are actually made, sooner or later, with probability 1.[22] It is an interesting question (discussed in Bostrom [2002b]), then, how to make sense of such a Big World cosmology as having observational consequences *at all*. The self-sampling assumption is a principle that can solve this problem. The idea is that our main datum is not that *someone* in the universe makes an observation, but that *we* make an observation. We have, then, an indifference principle with an essential *de se* component, from which we infer that whichever theory accommodates this *de se* datum better is the one with higher probability. For instance, the Big World hypothesis would have it that with probability one, sooner or later, an ordinary object such as a human brain will

[21] An imaginary device meant to detect conscious experience in others, invented by Chalmers, and presented, jokingly, in the guise of a hair-dryer by him during the second 'Toward a Science of Consciousness' conference, Tucson, Arizona, April 1996.

[22] For instance, if it is possible/conceivable that the above-mentioned fine-structure constant, α, has a different value than actually, that value would *actually* be instantiated with some nonzero probability, which means that it is not a constant after all. It is worth noting that, in fact, John K. Webb *et al.* ([1999], [2001]) have found evidence compatible with a slight time-variation of α with lower value in the past.

pop out of a black hole, because such phenomena are possible according to the theory. However, the fact that we do not observe such phenomena reduces their probability, virtually, to nil.

The reply to our objection could then be that the evaluation of evidence that disconfirms NNP being actually the case depends on whether such evidence is possible/conceivable; however, on the hypothesis of there being an NNP-conforming pattern of mental–physical correlation, assumed by the naturalistic dualist, such evidence should come out as impossible.

There is a further problem, however. SSA solves the problem of the Big World hypothesis as regards the meaningfulness of observational effects by being capable of assigning a probability of almost 1 to the theories that conform to what we observe. So why can't the naturalistic dualist assume a weaker thesis, coupled with the Big World hypothesis, namely, that if, *per impossibile*, we could observe other people's phenomenal experiences, we would observe NNP-conform correlations all the time. This is a weaker thesis because of the *de se* component, *we*, involved in the statements about observed data, which is compatible with some 'freakish observers' (Bostrom [2002b]) observing phenomena that are contrary to NNP, and which have, because of SSA, a vanishingly small probability. With such an approach the naturalistic dualist has a point in that it looks as though a small probability of observing phenomena that are contrary to NNP is *a fortiori* sufficient to falsify physicalism.

The answer to this problem is similar to the one I gave above; whether there actually are freakish observers is a matter of what is conceivable in this respect. But what is conceivable is not independent of what *we* (rather than the alleged freakish observers) observe, and so the verdict of the argument stands just as before. If, on the other hand, the thesis of nomological dangling for psycho-physical laws is assumed, then, as shown in the argument, the non-NNP scenarios will not only have probability zero, but won't count as possibilities at all, because if they did, they wouldn't be equiprobable with the NNP state, contrary to the indifference principle sanctioned by the thesis of nomological danglers.

> ***Objection 3:*** *The argument seems to massively overgeneralize to any case of apparent inductive or abductive knowledge.*[23]

Everyone allows that it is conceivable that the world has the same regularities in the past but different regularities in the future—and, of course, there are infinitely many ways that the future could be different. By my reasoning, the objection goes, we can't know that the future regularities obtain. But most people think we can know this. So the argument seems to prove inductive skepticism. Likewise for abductive skepticism and the like.

[23] An objection raised by David Chalmers in correspondence.

In reply I would like to point out that when it comes to induction there is a clear disanalogy between the past-to-future generalization and the potential direct generalization from one's own experience to psycho-physical laws holding across all conscious subjects in a world. Induction in general involves inference to 'if *p*, then probably *q*' from a large number of observed instantiations of *p* correlated with *q*. However, in the case of phenomenal properties the only case in which such inductive reasoning can proceed is the first-person case, and, indeed, I have assumed in my argument that there are such laws in the first-person case, based on observation of regularities of the form '*p* is always correlated with *q*', where '*p*' stands for neurophysiological properties and '*q*' for qualia. Part of the assumption could, of course, be that we can inductively generalize from past correlations in our own case, because we did observe the correlations in the past. But in the case of trying to directly generalize from own case correlations to laws that hold across the board in a world, we lack the observation of a large number of instantiations of *q* in a large number of subjects; given the explanatory irrelevance of experience (Chalmers), or the fact that psycho-physical laws are nomological danglers (Feigl), we can never have any acceptable evidential bases for a standard inductive inference. However, we can derive indirectly that the actual world has the required laws, namely, by appeal to the indifference principle; and, indeed, we did derive that the actual world is in NNP. So there is nothing in my argument to entail skepticism about induction in general. All the argument shows is that in the mental–physical case, unlike in physical–physical cases, for instance, there is no direct inductive generalization that is feasible; but this should not be a surprise, as it just follows from both Feigl's and Chalmers' observations about the special properties of qualia in the context of explanation.

Turning to abduction, or inference to the best explanation, again, I see no reason to think that the argument is committed to any kind of skepticism, to say nothing of the fact that part of the argument is actually based on such reasoning. Abductive reasoning involves the inference to '*p* explains *q*' as the result of *p* being the best explanation—in terms of simplicity, prior probability, and explanatory power—from among the possible explanations of *q*. In our case '*p*' is a replacement for the totality of laws that constitute NNP, while '*q*' stands for the psycho-physical correlations observed in my own case. According to the second part of argument, when we applied the self-sampling assumption (SSA) the best explanation of *q* is indeed *p*, because otherwise the fact that *q* holds would be a huge coincidence. This way we eliminated the 'anomic danglers' disjunct in proposition (C) as a priori improbable given (SSA). Then, abductive reasoning occurs once more, namely, when inferring that identity is a more economical way of explaining the resulting necessary nomic connections.

So, all in all, our argument does not imply any generalized skepticism about either induction or abduction.

> ***Objection 4:*** *The actual world having NNP is provable via 'fading qualia' and 'dancing qualia' type thought experiments.*

Chalmers ([1996], Chapter 5) offers arguments, based on the above-mentioned thought experiments, for what he calls 'the principle of organizational invariance', which is a law of nature stating that systems that share the same functional organization will instantiate the same phenomenal property patterns, regardless of neurophysiological (or any other physical) differences that the systems might be characterized by. In the fading qualia case, we suppose that our initially rich phenomenal experience gradually fades until it completely disappears, as a result of our brain cells being replaced by microchips. The question is whether the functional organization can stay constant. Chalmers argues that it must change as a consequence, given that we would observe and report these changes. To suppose otherwise would be to completely disconnect phenomenal experience from cognition, a very unnaturalistic scenario. The dancing qualia thought experiment involves a device that is implanted in one's brain that can switch between our natural neurophysiological basis for phenomenal experience and some alternative artificial basis, and qualia inversion is supposed to happen as a result of switching to the artificial basis. If the experience changes 'before my eyes', as Chalmers puts it, then it will have an effect on the functionally defined components of my cognitive system—I will recognize the change and report it. So there is no change in qualia without a change in functional organization.

In reply, I would like to point out two things. One is that Chalmers himself does not take these thought experiments as *proving* that the actual world is in state NNP. He explicitly states that:

> These arguments from thought experiments are only plausibility arguments, as always, but I think they have considerable force. To maintain the natural possibility of absent and inverted qualia in the face of these thought experiments requires accepting some implausible theses about the nature of conscious experience, and in particular between consciousness and cognition. Given certain natural assumptions about this relationship, the invariance principle is established as by far the most plausible hypothesis. ([1996], pp. 250–1)

In other words, what can be established by these thought experiments is not that these scenarios can't actually be the case, but that they are not naturally possible, *given that we know what the actual laws of nature are*, or that they are implausible *given that we know which laws of nature are actually plausible*. So these considerations leave our argument intact.

Second, and more importantly, as has recently been pointed out by Michael Pelczar ([2008]), Chalmers' thought experiments do not prove anything more than that within a *single* consciousness, across various temporal stages of it, there can't be phenomenal changes without functional changes. This has no effect on the *inter*personal case, when two distinct cognitive systems share their functional organization but they are inverted, or one of them has faded experiences. It is compatible with all we learn from the *intra*personal considerations about the fact that I would notice a change in my qualia, that there could theoretically be another person who has had her phenomenal experiences faded, as compared to mine, since birth. She would also notice changes *within her own* phenomenal field, but that does not change the fact that, as compared to mine, she has faded experiences. Even my own phenomenal experience might very well be faded, even when it is rich enough, with respect to some other person who, as a matter of fact, always has a much richer experience without functionally differing from me.

> ***Objection 5:*** *The argument seems to work against some epistemic arguments against physicalism, viz. the conceivability of zombies and that of qualia inversion, but when it comes to the knowledge argument it loses its attractiveness, as it entails that Mary, the superscientist who knows everything physical about the world but hasn't ever experienced any color, would not come to learn anything new when visually experiencing a red rose for the first time.*

The intuitive verdict in Mary's case is that she indeed learns something new when first seeing a red rose. If my argument is right, then what it is like to see red, for instance, is identical to a neurophysiological property, so she should not learn anything new as she knew all neurophysiological facts before having the experience of the red rose. However, our overall line of reasoning is not in conflict with the *prima facie* intuition that Mary learns something new, just as in the zombie case our line of thought starts from the supposition of the conceivability of them and of all the other non-NNP scenarios. What the argument shows is that the intuition is *ultimately* wrong, based on a priori probabilistic considerations. Of course, all this is consistent with the initial attractiveness of the intuition.

Let us see how exactly our line of reasoning applies to the knowledge argument. Remember that when we discussed the conceivability argument I pointed out that we've been conditioned to focus on certain rhetorically salient scenarios, like the zombie world and the inverted qualia world. In the case of the story of Mary we've been conditioned to focus on her coming to know what it is like to see red when first seeing a red rose. But why exactly *red*? Why, that is, do we think that Mary does indeed come to know what it is like to see red, rather than what it is like to see green, when first seeing a red rose?

David Lewis ([1990]) formulated, correctly, the knowledge argument via what he calls 'phenomenal information', namely, information containing possibilities that are left open by Mary's complete physical knowledge. According to this formulation, before seeing the red rose, Mary's complete physical knowledge leaves open infinitely many phenomenal possibilities about the world. Seeing the red rose for the first time is equivalent to the elimination of all these possibilities but one. But, for all we know so far, we are not justified in thinking that the possibility that is actualized in Mary's phenomenal field is phenomenal red. On the contrary, given that phenomenal red is only one such possibility out of an infinity of phenomenal possibilities that are left open by Mary's complete physical knowledge, we should assign a very low probability to this proposition. However, from my own case, that is, from the fact that I do experience what it is like to see red when seeing a red rose, together with the indifference principle, I can probabilistically infer that the actual world is in NNP, so Mary does experience what it is like to see red when seeing a red rose, rather than what it is like to see green or any other color. But if this is so, then almost surely in all possible worlds that are physical duplicates of the actual one Mary comes to know what it is like to see red, and not any other color, when seeing the red rose for the first time, otherwise the actual world being in NNP would be a huge coincidence.

This last proposition is equivalent to the proposition that the relation between Mary's brain state when seeing the red rose and what it is like to see red is epistemically necessary, as we arrived to this proposition by a priori probabilistic reasoning. Since this necessity should not be accepted as brute, we can posit the relation of identity between Mary's type of brain state and what it is like to see red, as that would explain why the correlation holds of necessity. This means that our initial intuition that Mary does learn something new when seeing red for the first time is ultimately mistaken, but, of course, as we proceeded through the steps of our probabilistic argument under the assumption—for the sake of a probabilistic *reductio*—of phenomenal information that is supposed to eliminate possibilities that are left open by physical knowledge, our line of reasoning is perfectly compatible with the *existence* of the intuition that Mary learns something new, but, of course, ultimately incompatible with its *truth*.

The conclusion is as radical as Lewis's own conclusion, i.e. that there is no such thing as phenomenal information, or as Daniel Dennett's ([1991]) reply to the knowledge argument, according to which Mary simply does not learn anything new. The difference is that, in my view, neither Lewis nor Dennett made a good enough case for the conclusion, beyond asserting it. The argument offered here looks at least to be one good candidate for the a priori *derivation*, via plausible probabilistic principles, of this radical conclusion.

6 Objections Related to Technicalities

Objection 6: *The indifference principle is known to lead to inconsistencies.*

Indeed, there is a lively discussion about how to formulate indifference principles so as to avoid inconsistent probability assignments. We can generate inconsistent probability assignments by coarsening the outcome space. To take a simple example, consider that all we know is that there are three buckets and one of them contains a ball, but we can't see the contents of the buckets. We are required to assign a probability distribution of the ball being present in a bucket over the three buckets. The indifference principle tells us that since there is no reason to prefer one bucket over any other when it comes to guessing whether the ball is present in them, we should assign equal probability for each bucket to contain the ball, which is 1/3. So, for instance, supposing we name the buckets as B_1, B_2, and B_3, the probability of the ball being in B_1 is 1/3. We can now coarsen the outcome space by renaming some of the outcomes, for instance, as:

Outcome 1: ball is in B_1.

Outcome 2: ball is in B_2-or-B_3.

Since the number of outcomes now is 2, the indifference principle will prescribe a probability of 1/2 for the ball being in B_1, which is inconsistent with the previous assignment. Yet, we used the very same principle of indifference.

As applied to our problem, we could coarsen the space of conceivable worlds that are physical duplicates of actuality but with different distributions of phenomenal properties, by renaming them, as follows:

Outcome 1: the actual world is in NNP.

Outcome 2: the actual world is in non-NNP.

Again, since we have two outcomes the indifference principle will sanction a probability of 1/2 for each, which means that, contrary to what we've been arguing for, we should suspend judgment about whether naturalistic dualism is true or the disjunction between anomic danglers and necessary danglers.

In reply, one could argue that the notion of a possible world as a maximal consistent set of propositions is clear enough to exclude disjunctive coarsening of the outcome space. A phrase like 'possible world *W*-or-*W**' does not refer to a possible world at all if *W* and *W** are themselves maximal consistent sets of propositions, whereas a disjunction of the form 'possible world *W* or possible world *W**' will always refer to either of the two worlds but never to both or to some fusion of them.

At the same time, we can also appeal to some consistent restriction of the indifference principle. Paul Castell ([1998]) offers such a restriction, which he calls 'the irrelevance principle'. Instead of considering the number

of outcomes in the outcome space and assigning equal probability to these, we consider a physical system, *P*, and one particular outcome, *O*, that the system can be in; after this we assert that that the probability of *P* being in *O* is the same at all times, or that the probability of each duplicate of *P* being in *O*, given some time, is constant. We then repeat the same reasoning with respect all the other outcomes besides *O*. The probability of a particular outcome will be given by the frequency of truth of the proposition stating the outcome within the set of propositions describing each of the duplicates of the system, or the system itself at different times.

To exemplify, consider our 'ball and buckets' example. The system is the ball and the bucket, and buckets 1, 2, and 3 are assumed to be duplicates. The relevant state or outcome is the ball being present in the bucket, which we will denote by '1' (the state of the ball being absent will be denoted by '0'). We can represent the problem as follows:

A_1: system 1 (i.e. B_1) is in State 1.

A_2: system 2 (i.e. B_2) is in State 1.

A_3: system 3 (i.e. B_3) is in State 1.

What the irrelevance principle sanctions is that propositions A_1–A_3 are equiprobable. The particular number is then given by assigning TRUE to a proposition of the form 'system *x*, for some *x*, is in State 1', and observing the frequency of truth about the system being State 1 in the set of propositions A_1–A_3:

TRUE: 'system 1 (i.e. B_1) is in State 1'.

FALSE: 'system 2 (i.e. B_2) is in State 1'.

FALSE: 'system 3 (i.e. B_3) is in State 1'.

That is, according to our problem, whenever one of the systems is in State 1, the other systems must be in State 0. Hence, we obtain the probabilities 1/3 for State 1 and 2/3 for State 0, for a particular system *x*. The problem of inconsistent probability assignments via disjunctive redescription of the outcome space is solved because our above-mentioned Outcome 2 (i.e. ball is in B_2-or-B_3) is not itself a duplicate of our physical system.

Applying this reasoning to our problem, we consider as our physical system the totality of physical facts of the actual world, call it φ, and the relevant state as NNP. φ will have an infinity of duplicates, if the first premise of the conceivability argument and PSYCHO-EXPLOSION are true, each duplicate corresponding to a rearrangement of phenomenal properties. Then we can describe our problem as:

A_1: system φ_1 is in state NNP.

A_2: system φ_2 is in state NNP.

. . .

A_n: system φ_n is in state NNP.

Propositions A_1–A_n will be equiprobable. Further, the probability of the system being in NNP will always be $1/n$, whereas the probability of the system being in non-NNP will always be $(n-1)/n$, because assigning TRUE to A_1 renders A_2–A_n all false. A redescription of the form 'system φ_1 or φ_2' won't be allowed as it would refer to a system that is not a duplicate of φ, i.e. not a physical duplicate of the actual world.

Finally, since n is a very large number, the probability of the actual world being in state NNP is virtually zero. Hence, the conclusion that either the conceivability premise is false, or it is almost sure that we don't live in a psycho-physically nomological world.

> ***Objection 7:*** *The indifference principle applies when the physical systems required for stating the propositions of the outcome space actually exist; but conceivable worlds do not exist, so the principle is not applicable.*

Of course, many times these systems actually exist. For instance, in our 'ball and buckets' example all the buckets exist. Similarly, consider the problem of assigning a probability to a particular poker card being an Ace of Clubs, when ignorant about any other fact about the cards. We assign probability 1/52, and the other 51 cards of the deck exist. But what is important is not whether or not the physical systems that carry the unactualized states exist, but only the conceptual possibility of these systems, that is, the existence of an abstract representation of all these systems. In the card game example, we would obtain the very same result, had all the other cards been destroyed, except the one we are presented with. We can even imagine God creating a universe with only one poker card, with the same results of the application of the indifference principle. All we need is an abstract representation of the game of poker as containing 52 cards, the Ace of Clubs being among them.[24]

[24] Sometimes the indifference principle is used as a way to argue for the existence of the physical systems that support the nonactualized possibilities. The argument for the existence of the Multiverse is such an example. Here, the variable is whether the universe contains life with conscious observers. The multiverse theorist argues as follows. Given (a) the fine-tuning of our universe (i.e. the extreme sensitivity of our variable to the physical magnitudes and constants of the initial conditions), and (b) the fact that we do live in such a universe, we would be either completely unsurprised, had our universe been just brutely there, or extremely surprised, had the magnitudes and constants of this universe been probabilistically 'selected', given that the universe containing conscious observers is one case in a very large number of possible lifeless universes. But given (a) we should not be completely unsurprised, and given (b) we should not be extremely surprised either. The only way to find a moderate level of our surprise is, therefore, to assume the existence of a multitude of universes, most of them characterized by all the non-actual values of the magnitudes and constants of the initial conditions, and one of them being our universe. Given all these universes, it is no surprise that one of them contains life, but it is still somewhat surprising as the frequency of life-containing universes within the multiverse is extremely low. The interested reader might consult John Leslie's ([1989]).

7 Conclusion

We started out with the early mind–brain identity thesis, and after a detour through the dialectic that followed as regards the mind–body problem, we reached the same conclusion that Feigl, Smart, and Place argued for, but in a more roundabout way, taking into account the strongest arguments for naturalistic dualism. Feigl's notion of a nomological dangler and its implications helped us build a probabilistic argument against merely nomological danglers, and opened the way to the final step, that of identifying mental and physical properties. If the argument is judged to have any attractiveness to it, it should be treated as a new challenge to dualists.

Acknowledgements

I am grateful for feedback on an earlier version from David Chalmers and the audience at Middle East Technical University, Ankara. Many thanks go to several referees for *BJPS* whose comments on the penultimate version improved the paper considerably. I would like to also thank TÜBITAK, *The Scientific and Technological Research Council of Turkey*, for continued support of my research.

Department of Philosophy
Bilkent University
Bilkent, Ankara, 06800
Turkey
istvanaranyosi@gmail.com;
aranyosi@bilkent.edu.tr

References

Aranyosi, I. [2008]: 'Excluding Exclusion: The Natural(istic) Dualist Approach', *Philosophical Explorations*, **11**, pp. 67–78.

Armstrong, D. M. [1968]: *A Materialist Theory of the Mind*, London: Routledge.

Bostrom, N. [2002a]: *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, New York: Routledge.

However, the multiverse case is very different from the case that supports our argument. In the first application of the indifference principle, i.e. when applied to the actual world considered among the set of all physical duplicate worlds, condition (b) is not satisfied, as *ex hypothesi* we do not, given the notion of a nomological dangler, observe the phenomenal property instantiations of the actual world. In the second application of the indifference principle, i.e. when applied to own case phenomenal property instantiations considered among all such instantiations in the actual world, while (b) is satisfied, as I do observe my own phenomenal property instantiations, condition (a) is not satisfied, as there is no reason to think that there is any dependence of phenomenal property instantiations in the actual world on any own case physical particularities.

Bostrom, N. [2002b]: 'Self-Locating Belief in Big Worlds: Cosmology's Missing Link to Observation', *Journal of Philosophy*, **99**, pp. 607–23.

Castell, P. [1998]: 'A Consistent Restriction of the Principle of Indifference', *British Journal for the Philosophy of Science*, **49**, pp. 387–96.

Chalmers, D. J. [1995]: 'Facing Up to the Problem of Consciousness', *Journal of Consciousness Studies*, **2**, pp. 200–19; Reprinted in S. Hameroff, A. Kaszniak and A. Scott (eds) [1996]: *Toward a Science of Consciousness*, Cambridge, MA, MIT Press, pp. 5–28.

Chalmers, D. J. [1996]: *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Chalmers, D. J. [2003]: 'Consciousness and Its Place in Nature', in S. P. Stich and T. A. Warfield (*eds*), *The Blackwell Guide to Philosophy of Mind*, Oxford: Blackwell Publishing, pp. 102–42.

Crane, T. [2001]: 'The Significance of Emergence', in C. Gillet and B. Loewer (*eds*), *Physicalism and its Discontents*, Cambridge: Cambridge University Press, pp. 207–24.

Dennett, D. [1991]: *Consciousness Explained*, Boston: Little Brown and Company.

Elga, A. [2004]: 'Infinitesimal Chances and the Laws of Nature', *Australasian Journal of Philosophy*, **82**, pp. 67–76.

Feigl, H. [1958/1967]: 'The "Mental" and the "Physical" ', in H. Feigl, M. Scriven, and G. Maxwell (*eds*), 1958, *Concepts, Theories and the Mind-Body Problem*, Minnesota Studies in the Philosophy of Science, Vol. 2, Minneapolis: University of Minnesota Press, pp. 370–497. Reprinted with a Postscript in Feigl, H., 1967, *The 'Mental' and the 'Physical', The Essay and a Postscript*, Minneapolis: University of Minnesota Press.

Herzberg, F. [2007]: 'Internal laws of probability, generalized likelihoods and Lewis' infinitesimal chances–A response to Adam Elga', *British Journal for the Philosophy of Science*, **58**, pp. 25–43.

Jackson, F. C. [1982]: 'Epiphenomenal Qualia', *Philosophical Quarterly*, **32**, pp. 127–36.

Keisler, J. H. [2000]: *Elementary Calculus: An Infinitesimal Approach*, Online edition: <www.math.wisc.edu/~keisler/calc.html>.

Kirk, R. [1974a]: 'Sentience and Behaviour', *Mind*, **83**, pp. 43–60.

Kirk, R. [1974b]: 'Zombies v. Materialists', *Proceedings of the Aristotelian Society*, **48**, pp. 135–52.

Kripke, S. A. [1972]: *Naming and Necessity*, Cambridge, MA: Harvard University Press.

Leslie, J. [1989]: *Universes*, New York: Routledge.

Levine, J. [1983]: 'Materialism and Qualia: The Explanatory Gap', *Pacific Philosophical Quarterly*, **64**, pp. 354–61.

Lewis, D. K. [1966]: 'An Argument for the Identity Theory', *Journal of Philosophy*, **63**, pp. 17–25.

Lewis, D. K. [1970]: 'How to Define Theoretical Terms', *Journal of Philosophy*, **67**, pp. 427–46.

Lewis, D. K. [1972]: 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, **50**, pp. 249–58.

Lewis, D. K. [1988]: 'Relevant Implication', *Theoria*, **54**, pp. 161–74. Reprinted in Lewis, D. K. [1998]: *Papers in Philosophical Logic*, Cambridge Studies in Philosophy, Cambridge: Cambridge University Press, pp. 111–24.

Lewis, D. K. [1990]: 'What Experience Teaches', in W. G. Lycan (*ed.*), *Mind and Cognition: A Reader*, Oxford: Blackwell, pp. 499–519. Reprinted in Lewis, D.K. [1999]: *Papers in Metaphysics and Epistemology*, Vol. 2. Cambridge: Cambridge University Press, pp. 262–91.

McCall, S. and Armstrong, D. M. [1989]: 'God's Lottery', *Analysis*, **49**, pp. 223–4.

McGinn, C. [2001]: 'How Not to Solve the Mind-Body Problem', in C. Gillett and B. M. Loewer (*eds*), *Physicalism and its Discontents*, Cambridge: Cambridge University Press, pp. 284–306.

Nagel, T. [1974]: 'What is it Like to Be a Bat?', *Philosophical Review*, **83**, pp. 435–50. Reprinted in Nagel, T. [1979]: *Mortal Questions*, Cambridge: Cambridge University Press, pp. 165–80.

Nolan, D. [1997]: 'Impossible Worlds: A Modest Approach', *Notre Dame Journal for Formal Logic*, **38**, pp. 535–72.

Papineau, D. [2002]: *Thinking about Consciousness*, Oxford: Oxford University Press.

Pelczar, M. [2008]: 'On an Argument for Functional Invariance', *Minds and Machines*, **18**, pp. 373–7.

Place, U. T. [1956]: 'Is Consciousness a Brain Process?', *British Journal of Psychology*, **47**, pp. 44–50.

Skyrms, B. [1995]: 'Strict Coherence, Sigma Coherence, and the Metaphysics of Quantity', *Philosophical Studies*, **77**, pp. 39–55.

Smart, J. J. C. [1959]: 'Sensations and Brain Processes', *Philosophical Review*, **68**, pp. 141–56.

Stoljar, D. [2000]: 'Physicalism and the Necessary A Posteriori', *Journal of Philosophy*, **97**, pp. 33–54.

Stoljar, D. [2001]: 'Physicalism', in E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy*; <plato.stanford.edu/entries/physicalism/>.

Strawson, G. [1997]: 'The Self', *Journal of Consciousness Studies*, **4**, pp. 405–28.

Stubenberg, L. [1998]: *Consciousness and Qualia*, Amsterdam: John Benjamins.

Vallentyne, P. [2000]: 'Standard Decision Theory Corrected', *Synthese*, **122**, pp. 261–90.

Webb, J. K., Flambaum, V. V., Churchill, C. W., Drinkwater, M. J. and Barrow, J. D. [1999]: 'Search for Time Variation of the Fine Structure Constant', *Physical Review Letters*, **82**, pp. 884–7.

Webb, J. K., Murphy, M. T., Flambaum, V. V., Dzuba, V. A., Barrow, J. D., Churchill, C. W., Prochaska, J. X. and Wolfe, A. M. [2001]: 'Further Evidence for Cosmological Evolution of the Fine Structure Constant', *Physical Review Letters*, **87**, 091301.

Williamson, T. [2007]: 'How Probable is an Infinite Sequence of Heads?', *Analysis*, **67**, pp. 173–80.